BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: WO 00/58516

(43) International Publication Date: 5 October 2000 (05.10.00)

(21) International Application Number: PCT/US00/08069

(22) International Filing Date: 27 March 2000 (27.03.00)

(30) Priority Data:

60/126,473	26 March 1999 (26.03.99)	US
60/140,359	23 June 1999 (23.06.99)	US

(71) Applicants: WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors: FAN, Jian-Bing; 3737 Nobel Drive, #2115, San Diego, CA 92122 (US). HIRSCHHORN, Joel, N.; 14 Oakmont Road, Newton, MA 02459 (US). HUANG, Xiaohua; 937 Jackson Street, Mountain View, CA 94043 (US). KAPLAN, Paul; 2190 Vizcaya Circle, Campbell, CA 95008 (US). LANDER, Eric, S.; 151 Bishop Allan Drive, Cambridge, MA 02139 (US). LOCKHART, David, J.; 510 Torrey Point Road, Del Mar, CA 92014 (US). RYDER, Thomas; 100 Spring Street, Los Gatos, CA 95030 (US). SKLAR, Pamela; 68 Fuller Street, Brookline, MA 02446 (US).

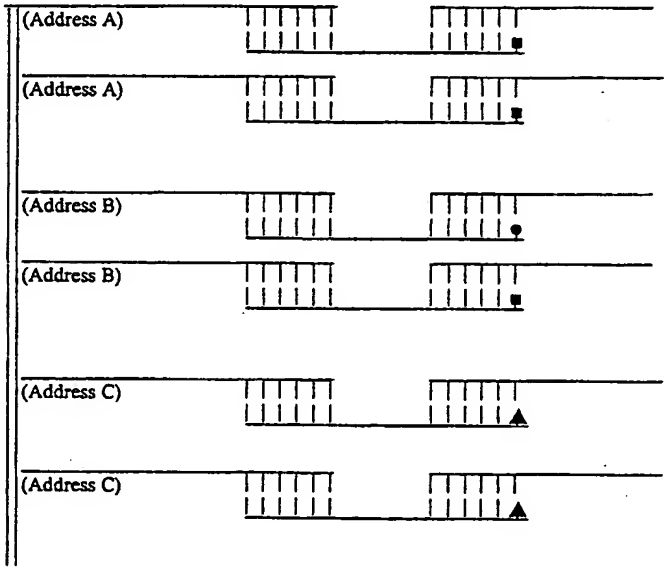
(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: UNIVERSAL ARRAYS



(57) Abstract

An array of oligonucleotides on a solid substrate is disclosed, which can be used for multiple purposes. Methods and reagents are provided for performing genotyping to determine the identity or ration of allelic forms of a gene in a sample. A single base extension primer is coupled to a sequence identity code. During the primer extension reaction a distinctive label is incorporated which identifies the allelic form present in the sample. This permits multiple simultaneous analyses to be performed easily and efficiently.

UNIVERSAL ARRAYS

BACKGROUND OF THE INVENTION

Obtaining genotype information on thousands of polymorphic markers in a highly parallel fashion is becoming an increasingly important task in mapping disease loci, in identifying quantitative trait loci, in diagnosing tumor loss of heterozygosity, and in performing linkage studies. A currently available method for simultaneously obtaining large numbers of polymorphic marker genotypes involves hybridization to allele specific probes on high density oligonucleotide arrays. In order to practice the method, redundant sets of hybridization probes, typically twenty or more, are used to score each marker. A high degree of redundancy is required, however, to reduce the noise and achieve an acceptable level of accuracy. Even this level of redundancy is often insufficient to unambiguously score heterozygotes or to quantitatively determine allele frequency in a population. Thus, there is a need in the art for more reliable and better quantitative methods to identify genotypes at polymorphic markers.

The invention further relates to a method of genotyping a nucleic acid sample at one or more loci, comprising the steps of obtaining a nucleic acid sample to be tested; combining the nucleic acid sample with one or more locus-specific tagged oligonucleotides under conditions suitable for hybridization of the nucleic acid sample to one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex; subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs; contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementarysequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags, thereby determining the genotype of the queried nucleotide in the sample. In one embodiment the nucleic acid sample to be tested is amplified.

In one embodiment a method is provided to aid in determining a ratio of alleles at a polymorphic locus in a sample. A pair of primers is used to amplify a region of a nucleic acid in a sample. In one embodiment, the region comprises a polymorphic locus, and an amplified nucleic acid product is formed which comprises the polymorphic locus. The amplified nucleic acid product is used as a template in a single base extension reaction with an extension primer, forming a labeled extension primer. The extension primer (also called a locus-specific tagged oligonucleotide herein)

hybridized to one or more probes which are immobilized to known locations on a solid support.

These and other embodiments of the invention which are described in more detail below provide the art with methods and tools for rapidly and easily determining genotypes of individuals and allele frequencies in populations.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram of the universal array. The solid substrate (*e.g.*, a glass slide) is depicted on the left, and different oligonucleotide tags ("A", "B", "C", etc.) are shown attached to the solid substrate. The nucleotide sequence on the right-hand end of each oligonucleotide tag ("Tag A", Tag B", "Tag C") is arbitrary unique sequence; that is, it is designed and synthesized to be unique to each oligonucleotide tag.

Fig. 2 is a diagram depicting a locus-specific tagged oligonucleotide. The nucleotide sequence at the left-hand end is complementary to the arbitrary sequence of one of the oligonucleotide tags depicted in Fig. 1. The nucleotide sequence at the right-hand end is complementary to the amplification product of a known polymorphic locus (*e.g.*, a single nucleotide polymorphism (SNP)). Therefore, locus-specific tagged oligonucleotide "A" comprises anucleotide sequence complementary to the arbitrary sequence of the "Tag A" oligonucleotide tag depicted in Fig. 1, and also comprises sequence complementary to SNP "A".

Fig. 3 is a diagram showing the hybridization of the locus-specific tagged oligonucleotide to the amplification product. The locus-specific sequence (right hand end) of the oligonucleotide is designed so that it terminates one nucleotide immediately before (5' of) the nucleotide to be genotyped (shown in box).

Fig. 4 is a diagram depicting the labeling of the locus-specific tagged oligonucleotide-amplification primer complex via single base extension. During the reaction, a single labeled ddNTP complementary to the queried nucleotide is enzymatically added to the 3' end of the locus-specific tagged oligonucleotide. The nucleotide is shown in the box.

5'-GAACGCAGTTATCAGACTCTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 6);

5'-CGAGGACATGGAGTCACATCCAGGATCTTTCAGGTAGC-ACT-3' (SEQ ID NO: 7); and

5'-GCTAGGCATTCCTCCAGTGTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 8)) were separately added to six SBE reactions which contain the mixed templates of different ratios. The SBE primers were extended in the presence of biotin-labeled ddATP and fluorescein-labeled ddCTP (see Examples) and pooled and hybridized to the tag array. The intensity ratio of the two colors (the y-axis) were plotted against the ratio of the mixed two templates (the x-axis).

Fig. 9 shows a clustering analysis of the tag array hybridization results in 44 individuals at marker GMP-140.25.

DETAILED DESCRIPTION OF THE INVENTION

The invention features a generic or universal genotyping array, consisting of oligonucleotide tags attached to a solid substrate (Fig. 1). Each address in the array (*e.g.*, "A", "B", "C", etc.) has an oligonucleotide tag associated with it. The oligonucleotide tag at a given address is attached to the solid substrate, and comprises a unique arbitrary nucleotide sequence. That is, the nucleotide sequence is unique for the oligonucleotide tag at each address, i.e., the nucleotide sequence for "tag A" is different from the nucleotide sequence for all other tags in the array. The nucleotide sequence for each tag is arbitrary in that it can be any sequence, provided that it is different from the nucleotide sequence for every other tag in the array. Preferably the oligonucleotide tag is from about 20 to about 50 nucleotides in length. It may also be desirable to design the nucleotide sequence of the oligonucleotide tag such that it does not facilitate an undesirable interaction, e.g., with the target nucleic acid molecule (amplified product).

The oligonucleotide array is used in conjunction with locus-specific tagged oligonucleotides. Each oligonucleotide tag in the array corresponds to a locus-specific tagged oligonucleotide. One end (the 5' end) of the locus-specific tagged

After the single base extension reaction, the complex of the labeled (extended) locus-specific tagged oligonucleotide and the amplification product is hybridized to the array (Fig. 5). The oligonucleotide tag "A" at address "A" selectively hybridizes to its corresponding locus-specific tagged oligonucleotide (now extended with a labeled ddNTP), the oligonucleotide tag "B" at address "B" selectively hybridizes to its corresponding locus-specific tagged oligonucleotide (now extended with a labeled ddNTP), etc. The array is assayed to determine which label(s) is (are)present at which address on the array. For instance, if address "A" fluoresced at the same wavelength as the label on the ddATP, then the amplification product clearly contained a "T" at the queried nucleotide (because the single base extension reaction attaches the ddNTP complementary to the queried nucleotide). Fluorescence at a wavelength which is the same as the ddCTP label would indicate that the genotype was a "G", etc. Detection of two peaks within the wavelength emitted would indicate that different nucleotides were present at the queried position in the sample, e.g., that the individual was heterozygous at that locus.

An advantage of the array and method described herein is that many addresses can be assayed simultaneously, producing genotyping data for many different genetic loci, e.g., SNPs. By utilizing a predefined set of locus-specific tagged oligonucleotides, *e.g.*, a set specific for assaying a set of genetic diseases, a single array can be utilized for a particular purpose, and by utilizing a different set of locus-specific tagged oligonucleotides which correspond to the same tags on the array, the same array can be utilized for a different purpose. The universal chip serves as the repository of a set of addresses to which the locus-specific tagged oligonucleotides (along with the labeled, genotyped SNPs) hybridize in a planned, predetermined manner. The array and set(s) of locus-specific tagged oligonucleotides can therefore be used as components in kits for the purposes of sequencing and genotyping. Sets of locus-specific tagged oligonucleotides can therefore be used in combination with arrays as described herein for use in forensics, identification of individuals, and disease diagnosis/prognosis.

more than 1 kb, 0.5 kb, 0.2 kb, 0.1 kb, 0.01 kb or 0.001 kb apart. A suitable DNA polymerase can be used as is known in the art. Thermostable polymerases are particularly convenient for thermal cycling of rounds of primer hybridization, polymerization, and melting. Amplification of single stranded nucleic acids can also be employed.

After the amplification it is desirable to remove and/or degrade any excess primers and nucleotides. This can be done by washing and/or enzymatic degradation, using such enzymes as endonuclease I and alkaline phosphatase, for example. Other techniques, such as chromatography, magnetic beads, and avidin- or streptavidin-conjugated beads, as are known in the art for accomplishing the removal can also be used. It is not necessary to remove or destroy one of two strands of an amplified DNA product.

The primer extension step of the method is the one which provides allele-specificity to the method. The primer is designed to terminate one nucleotide 5' to the polymorphic locus. The primer is hybridized to the denatured amplified double stranded DNA. When the primer is extended by a single base using dideoxynucleotides and a DNA polymerase, the dideoxynucleotide which is complementary to the nucleotide at the polymorphic locus is added. Again, any DNA-dependent DNA polymerase can be used. These include, but are not limited to, *E. coli* DNA polymerase I, Klenow fragment of polymerase I, T4 DNA polymerase, T7 DNA polymerase, *T. aquaticus* DNA polymerase. This reaction is preferably performed at the T_M of the primer with the template to enhance product formation.

One configuration for carrying out the primer extension step utilizes two different primers which each hybridize to opposite strands of an amplified double stranded DNA. Each primer terminates one nucleotide 5' to the polymorphic locus. The primer extension reaction may be more robust with one strand as a template than the other. In addition, the information obtained from the second strand should confirm the information obtained from the first strand.

The labels which are used can be any which are known in the art. These include radiolabels, fluorescent labels, enzyme labels, epitope labels, and high affinity binding partner labels. Examples include isotopically labeled nucleotides, fluorescein-labeled nucleotides, biotin-labeled nucleotides, digoxin labeled nucleotides. A different label is assigned to each base dideoxynucleotide in the single base extension reaction. Two, three, or four different labels can be used in the reaction. The different labels can be all of the same type, *e.g.*, enzyme labels, or they can be mixed types.

Hybridization of the 5' portion of the extension primers (the tag sequences) to one or more probes which are immobilized to known locations on a solid support is also contemplated. Hybridization can be performed under standard conditions known in the art for obtaining robust signals at high specificity. Standard washing conditions can also be employed. Detection of hybridization of the extension primers can be done using standard means, depending on the type of labels used. For example, fluorescence can be detected and quantified using optical detection means. Radiolabels can be detected using autoradiography or scintillation counting. Enzyme labels can be detected using enzymatic reactions and assaying for the final product of the enzyme reaction. Antigenic labels can be used using immunological detection means. Affinity binding partners such as strepavidin or avidin and biotin can also be used as a label.

The reactions of the present invention can be performed in a single or multiplex format. For example, the amplification step can be performed using up to 20, 30, 40, 50, 75, 100, 150, 200, 250, or 300 different primer pairs to amplify a corresponding number of polymorphic markers. These can be pooled for the single base extension reaction, if desired. Pooling for the hybridization step is desirable so that thousands of hybridizations can be done simultaneously.

In an alternative embodiment the amplification step can be omitted. Thus, if sufficient DNA is available, the single base extension reaction can be performed directly on genomic DNA. In another particular embodiment, amplification of the entire genome can be performed using random primers.

preferred that the array include one or more control probes. In one embodiment, the array is a high density array. A high density array is an array used to hybridize with a target nucleic acid sample to detect the presence of a large number of allelic markers, preferably more than 10, more preferably more than 100, and most preferably more than 1000 allelic markers.

High density arrays are suitable for quantifying small variations in the frequency of an allelic marker in the presence of a large population of heterogeneous nucleic acids. Such high density arrays can be fabricated either by *de novo* synthesis on a substrate or by spotting or transporting nucleic acid sequences onto specific locations of a substrate. Both of these methods produce nucleic acids which are immobilized on the array at particular locations. Nucleic acids can be purified and/or isolated from biological materials, such as a bacterial plasmid containing a cloned segment of a sequence of interest. Suitable nucleic acids can also be produced by amplification of templates or by synthesis. As a nonlimiting illustration, polymerase chain reaction and/or *in vitro* transcription, are suitable nucleic acid amplification methods.

The term "target nucleic acid" refers to a nucleic acid (either synthetic or derived from a biological sample or nucleic acid sample), to which the probe is designed to specifically hybridize. In this invention, such target nucleic acids are the same as the sequence tags. It is either the presence or absence of the target nucleic acid that is to be detected, or the amount of the target nucleic acid that is to be quantified. The target nucleic acid has a sequence that is complementary to the nucleic acid sequence of the corresponding probe directed to the target. The term "target nucleic acid" can refer to the specific subsequence of a larger nucleic acid to which the probe is directed or to the overall sequence (*e.g.*, gene or mRNA) whose presence it is desired to detect. The difference in usage will be apparent from context.

As used herein a "probe" is defined as a nucleic acid, capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe can include natural (*i.e.* A, G, U, C, or T) or

In addition to test probes that bind the target nucleic acid(s) of interest, the high density array can contain a number of control probes. The control probes fall into two categories: normalization controls and mismatch controls.

Normalization controls are oligonucleotide or other nucleic acid probes that are complementary to labeled reference oligonucleotides or other nucleic acid sequences that are added to the nucleic acid sample. The signals obtained from the normalization controls after hybridization provide a control for variations in hybridization conditions, label intensity, "reading" efficiency, and other factors that may cause the signal of a perfect hybridization to vary between arrays. In a preferred embodiment, signals (*e.g.*, fluorescence intensity) read from all other probes in the array are divided by the signal (*e.g.*, fluorescence intensity) from the control probes, thereby normalizing the measurements.

Virtually any probe can serve as a normalization control. However, it is recognized that hybridization efficiency varies with base composition and probe length. Preferred normalization probes are selected to reflect the average length of the other probes present in the array; however, they can be selected to cover a range of lengths. The normalization control(s) can also be selected to reflect the (average) base composition of the other probes in the array; however in a preferred embodiment, only one or a few normalization probes are used and they are selected such that they hybridize well (*i.e.* no secondary structure) and do not match any target-specific probes.

Mismatch controls can also be provided for the probes to the target alleles or for normalization controls. The terms "mismatch control" or "mismatch probe" or "mismatch control probe" refer to a probe whose sequence is deliberately selected not to be perfectly complementary to a particular target sequence. Mismatch controls are oligonucleotide probes or other nucleic acid probes identical to their corresponding test or control probes except for the presence of one or more mismatched bases. A mismatched base is a base selected so that it is not complementary to the corresponding base in the target sequence to which the probe would otherwise specifically hybridize. One or more mismatches are selected such that under appropriate hybridization

In a preferred embodiment, oligonucleotide probes in the high density array are selected to bind specifically to the nucleic acid target to which they are directed with minimal non-specific binding or cross-hybridization under the particular hybridization conditions utilized. Because the high density arrays of this invention can contain in excess of 100,000 or even 1,000,000 different probes, it is possible to provide every probe of a characteristic length that binds to a particular nucleic acid sequence.

Forming High Density Arrays

High density arrays are particularly useful for monitoring the presence of allelic markers. The fabrication and application of high density arrays in gene expression monitoring have been disclosed previously in, for example, WO 97/10365, WO 92/10588, U.S. Application Ser. No. 08/772,376 filed December 23, 1996; serial number 08/529,115 filed on September 15, 1995; serial number 08/168,904 filed December 15, 1993; serial number 07/624,114 filed on December 6, 1990, serial number 07/362,901 filed June 7, 1990, andin U.S. 5,677,195, all incorporated herein for all purposes by reference. In some embodiments using high density arrays, high density oligonucleotide arrays are synthesized using methods such as the Very Large Scale Immobilized Polymer Synthesis (VLSIPS) disclosed in U.S. Pat. No. 5,445,934 incorporated herein for all purposes by reference. Each oligonucleotide occupies a known location on a substrate. A nucleic acid target sample is hybridized with a high density array of oligonucleotides and then the amount of target nucleic acids hybridized to each probe in the array is quantified.

Synthesized oligonucleotide arrays are particularly preferred for this invention. Oligonucleotide arrays have numerous advantages over other methods, such as efficiency of production, reduced intra- and inter array variability, increased information content, and high signal-to-noise ratio.

Preferred high density arrays comprise greater than about 100, preferably greater than about 1000, more preferably greater than about 16.000, and most preferably greater than 65.000 or 250,000 or even greater than about 1,000.000 different oligonucleotide

reagent containing a functional group, *e.g.*, a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithogaphic mask is used selectively to expose functional groups which are then ready to react with incoming 5'-photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface. Combinatorial synthesis of different oligonucleotide analogues at different locations on the array is determined by the pattern of illumination during synthesis and the order of addition of coupling reagents.

In the event that an oligonucleotide analogue with a polyamide backbone is used in the VLSIPS™ procedure, it is generally inappropriate to use phosphoramidite chemistry to perform the synthetic steps, since the monomers do not attach to one another via a phosphate linkage. Instead, peptide synthetic methods are substituted. See, *e.g.*, Pirrung *et al.* U.S. Pat. No. 5,143,854.

Peptide nucleic acids are commercially available from, *e.g.*, Biosearch, Inc. (Bedford, MA) which comprise a polyamide backbone and the bases found in naturally occurring nucleosides. Peptide nucleic acids are capable of binding to nucleic acids with high specificity, and are considered "oligonucleotide analogues" for purposes of this disclosure.

Additional methods which can be used to generate an array of oligonucleotides on a single substrate are described in co-pending Applications Ser. No. 07/980,523, filed November 20, 1992, and 07/796,243, filed November 22, 1991 and in PCT Publication No. WO 93/09668. In the methods disclosed in these applications, reagents are delivered to the substrate by either (1) flowing within a channel defined on predefined regions or (2) "spotting" on predefined regions or (3) through the use of photoresist. However, other approaches, as well as combinations of spotting and flowing, can be employed. In each instance, certain activated regions of the substrate

regions are reacted with a monomer before the channel block must be moved or the substrate must be washed and/or reactivated. By making use of many or all of the available reaction regions simultaneously, the number of washing and activation steps can be minimized.

One of skill in the art will recognize that there are alternative methods of forming channels or otherwise protecting a portion of the surface of the substrate. For example, according to some embodiments, a protective coating such as a hydrophilic or hydrophobic coating (depending upon the nature of the solvent) is utilized over portions of the substrate to be protected, sometimes in combination with materials that facilitate wetting by the reactant solution in other regions. In this manner, the flowing solutions are further prevented from passing outside of their designated flow paths.

High density nucleic acid arrays can be fabricated by depositing presynthezied or natural nucleic acids in predetermined positions. Synthesized or natural nucleic acids are deposited on specific locations of a substrate by light directed targeting and oligonucleotide directed targeting. Nucleic acids can also be directed to specific locations in much the same manner as the flow channel methods. For example, a nucleic acid A can be delivered to and coupled with a first group of reaction regions which have been appropriately activated. Thereafter, a nucleic acid B can be delivered to and reacted with a second group of activated reaction regions. Nucleic acids are deposited in selected regions. Another embodiment uses a dispenser that moves from region to region to deposit nucleic acids in specific spots. Typical dispensers include a micropipette or capillary pin to deliver nucleic acid to the substrate and a robotic system to control the position of the micropipette with respect to the substrate. In other embodiments, the dispenser includes a series of tubes, a manifold, an array of pipettes or capillary pins, or the like so that various reagents can be delivered to the reaction regions simultaneously.

is performed at low stringency, in this case in 6X SSPE-T at 37°C (0.005% Triton X-100), to ensure hybridization. and then subsequent washes are performed at higher stringency (*e.g.*, 1 X SSPE-T at 37°C) to eliminate mismatched hybrid duplexes. Successive washes can be performed at increasingly higher stringency (*e.g.*, down to as low as 0.25 X SSPE-T at 37°C to 50°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity can be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present (*e.g.*, expression level control, normalization control, mismatch controls, *etc.*).

In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array can be washed at successively higher stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular oligonucleotide probes of interest.

The stability of duplexes formed between RNAs or DNAs are generally in the order of RNA:RNA > RNA:DNA > DNA:DNA, in solution. Long probes have better duplex stability with a target, but poorer mismatch discrimination than shorter probes (mismatch discrimination refers to the measured hybridization signal ratio between a perfect match probe and a single base mismatch probe). Shorter probes (*e.g.*, 8-mers) discriminate mismatches very well. but the overall duplex stability is low.

Altering the thermal stability (T_m) of the duplex formed between the target and the probe using, *e.g.*, known oligonucleotide analogues allows for optimization of duplex stability and mismatch discrimination. One useful aspect of altering the T_m arises from the fact that adenine-thymine (A-T) duplexes have a lower T_m than guanine-cytosine (G-C) duplexes. due in part to the fact that the A-T duplexes have two hydrogen bonds per base-pair. while the G-C duplexes have three hydrogen bonds per

preparation of the target nucleic acids. Thus, for example, polymerase chain reaction with labeled primers will provide a labeled amplification product.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (*e.g.*, Dynabeads™), fluorescent dyes (*e.g.*, fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horseradish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels can be detected using photographic film or scintillation counters, fluorescent markers can be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label. One method uses colloidal gold label that can be detected by measuring scattered light.

Means of detecting labeled target nucleic acids hybridized to the probes of the array are known to those of skill in the art. Thus, for example, where a colorimetric label is used, simple visualization of the label is sufficient. Where a radioactive labeled probe is used, detection of the radiation (*e.g.* with photographic film or a solid state detector) is sufficient.

Detection of target nucleic acids which are labeled with a fluorescent label (*i.e.*, a "color tag") can be accomplished with fluorescence microscopy. The hybridized array can be excited with a light source at the excitation wavelength of the particular fluorescent label and the resulting fluorescence at the emission wavelength is detected.

generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication, the frequency of an allele. Relative quantification can also be used to merely detect the presence or absence of an allele in the target nucleic acids. In one embodiment, for example, the presence or absence of the two alleles of a marker can be determined by comparing the quantities of the first and second color tag at the known locations in the array, *i.e.*, on the solid support, which correspond to the allele-specific probes for the two alleles.

A preferred quantifying method is to use a confocal microscope and fluorescent labels. The GeneChip® system (Affymetrix, Santa Clara, CA) is particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar system or other effectively equivalent detection method can also be used.

Methods for evaluating the hybridization results vary with the nature of the specific probes used, as well as the controls. Simple quantification of the fluorescence intensity for each probe can be determined. This can be accomplished simply by measuring signal strength at each location (representing a different probe) on the high density array (*e.g.*, where the label is a fluorescent label, detection of the florescence intensity produced by a fixed excitation illumination at each location on the array).

One of skill in the art, however, will appreciate that hybridization signals will vary in strength with efficiency of hybridization, the amount of label on the sample nucleic acid and the amount of the particular nucleic acid in the sample. Typically nucleic acids present at very low levels (*e.g.*, < 1 pM) will show a very weak signal. At some low level of concentration, the signal becomes virtually indistinguishable from background. In evaluating the hybridization data, a threshold intensity value can be selected below which a signal is counted as being essentially indistinguishable from background.

specific binding or the presence in the sample of a nucleic acid that hybridizes with the mismatch. Where both the probe in question and its corresponding mismatch control show high signals, or the mismatch shows a higher signal than its corresponding test probe, there is a problem with the hybridization and the signal from those probes is ignored. For a given marker, the difference in hybridization signal intensity ($I_{allele1}$ - $I_{allele2}$) between an allele-specific probe (perfect match probe) for a first allele and the corresponding probe for a second allele (or other mismatch control probe) is a measure of the presence of or concentration of the first allele. Thus, in a preferred embodiment, the signal of the mismatch probe is subtracted from the signal for its corresponding test probe to provide a measure of the signal due to specific binding of the test probe.

The concentration of a particular sequence can then be determined by measuring the signal intensity of each of the probes that bind specifically to that gene and normalizing to the normalization controls. Where the signal from the probes is greater than the mismatch, the mismatch is subtracted. Where the mismatch intensity is equal to or greater than its corresponding test probe, the signal is ignored (*i.e.*, the signal cannot be evaluated).

For each marker analyzed, the genotype can be unambiguously determined by comparing the hybridization patterns obtained for each of the two labels, *e.g.*, color tags employed (Fig. 8). If hybridization is indicated for one color tag to its corresponding allele-specific probe (*e.g.*, "A") but not for the other color tag (*e.g.*, "G") (pattern at left in Fig. 8), then the indicated genotype of a diploid organism would be homozygous A/A. If hybridization is indicated only for the other color tag to its corresponding allele-specific probe (*e.g.*, "G") (pattern at center in Fig. 8), then the indicated genotype of a diploid organism would be homozygous G/G. If hybridzation is indicated for both color tags to their corresponding allele-specific probes (pattern at right in Fig. 8), then the indicated genotype of a diploid organism would be heterozygoous (A/G).

Marginal detection of hybridization, indicated by an intermediate positive result (*e.g.*, less than 1%, or from 1-5%, or from 1-10%, or from 2-10%, or from 5-10%, or from 1-20%, or from 2-20%, or from 5-20%, or from 10-20% of the average of all

approach should be explored, for example. strategies involved using total human genomic DNA directly, or genomic DNA amplified using some general amplification methods, *e.g*,. primer-extension preamplification, PEP[25], or total cDNA. In fact, we have tried to use total human genomic DNA directly as the SBE template in our tag array assay. 24 out of the 38 of the markers that we tested gave good signals (data not shown). Nevertheless, large amount of work are warranted as to solve both the sensitivity (signal intensity) and specificity (mis-priming) problems before the whole-genome approach become really useful.

The invention will be further illustrated by the following non-limiting examples. The content of references cited herein is incorporated herein by reference in its entirety.

EXEMPLIFICATION

METHODS

Collection and Isolation of DNA From Samples

DNA samples were collected by GenNet as part of the ongoing Family Blood Pressure Program. Samples were collected with consent and IRB approval in both Tecumseh, MI and Loyola, IL FAMILIES. Ascertainment was based on identification of a proband in the top 15th (Tecumseh) or 20th (Loyola) percentile of the community's blood pressure distribution. Full phenotypic information was obtained for each individual. DNA was extracted from 5-10 ml of whole blood taken from each individual using the standard "salting-out" method (Gentra Systems).

Primer Design

For each SNP, primary PCR amplification primers were designed as described previously[9]. The SBE primer was designed in a manner that its 3' terminates one base before the polymorphic site. Primer 3.0 software package (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) was modified and used to pick SBE primers with batch sequences. at a predicted length of 20 (ranging from 18 to 26) nucleotide and melting temperature of 60°C (ranging from 54°C to 64°C). The SBE

Extension reaction was carried out on a Thermo Cycler (MJ Research), with 1 cycle of 96°C for 3 minutes, then 45 cycles of 94°C for 20 seconds and 58°C for 11 seconds.

After SBE reaction, 9 reactions from each sample were combined and mixed with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), 18.75 µl of 8 M LiCl (Sigma), and 1125 µl of pre-chilled (-20°C) ethanol (Abs.), and precipitated by centrifugation at the top speed (Eppendorf centrifuge 5415C) for 15 minutes at room temperature; precipitated samples were dried at 40°C for 40 minutes and re-suspended in 33 µl ddH_2O.

Tag Array Design and Hybridization

For each tag sequence, two probes were synthesized on the array. One is exactly the designed tag sequence (referred to as a Perfect Match, or PM probe). The other one is identical except for a single base difference in a central position (referred to as a Mismatch, or MM probe). The mismatch probe services as an internal control for hybridization specificity. Over 32,000 20-mer tag probes (and their companions) were chosen[11] and fabricated on a 8 mm x 8mm size of array. Each probe (feature) occupies a 30 microns x 30 microns area. The sets of arrays were synthesized together on a single glass wafer on which 100 arrays were made.

The labeled sample was denatured at 95°C - 100°C for 10 minutes and snap cooled on ice for 2 - 5 minutes. The tag array was pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5 mg/ml of BSA for a few minutes, then hybridized with 120 µl hybridization solution (as shown below) at 42°C for 2 hours on a rotisserie, at≅ 40 RPM. Hybridization Solution consists of 3M TMACL (Tetramethylammonium Chloride), 50 mM MES ((2-[N-Morpholino]ethanesulfonic acid) Sodium Salt) (pH 6.7), 0.01% of Triton X-100, 0.1 mg/ml of Herring Sperm DNA, 50 pM of fluorescein-labeled control oligo, 0.5 mg/ml of BSA (Sigma) and 29.4 µl labeled SBE products (see below) in a total of 120 µl reaction.

ABI Sequencing to Determine Genotypes

To independently confirm the genotypes called from the tag array assay, three samples (904957000000, 904896000000, and 904889000000) were sequenced using gel-electrophoresis based method. Samples were amplified for all sites with T7 and T3 tagged primers, using standard PCR cycling conditions (2.5 µl of 20 ng/µl DNA, 0.375 µl of 20 µM primer (X2), 1.5 µl of 10X PCR buffer, 0.9 µl 25mM Mg^{2+}, 0.15 µl 10mM dNTPs, 0.25 µl 10 U/µl Taq DNA Polymerase (Sigma), brought up to 15 µl with ddH_20 per tube). Some products were sequenced directly, while a M13 nesting strategy was used due to the close proximity of the polymorphic base to the primer end. Samples from the initial amplification were diluted 1:50 with ddH_20, and amplified with M13F-T7 (TGTAAAACGACGGCCAGTTAATACGACTCACTATAGGGAGA; SEQ ID NO: 9) and M13R-T3 (AACAGCTATGACCATGAATTAACCCTCACTAAAGGGAGA; SEQ ID NO: 10) primers using standard PCR conditions. All PCR products were cleaned with Exonuclease I (Amersham 0.15 µl of 10 U/µl per well) and Shrimp Alkaline Phosphatase (Amersham, 0.30 µl of 1 U/µl per well) in a volume of 10 µl. Dye terminator sequencing using a M13R primer (AACAGCTATGACCATG; SEQ ID NO: 11) or T7 primer (TAATACGACTCACTATAGGGAGA; SEQ ID NO: 12) on an ABI377 (Perkin Elmer) using Big Dyes (Perkin Elmer) was performed to determine the genotype status for each SNP in all three individuals. Trace files were read with Edit View 1.0 (Perkin Elmer) software.

EXAMPLE 1

DNA from a individual is isolated, and amplified with primers from 15 previously-characterized (i.e., known) SNPs. Amplification is allowed to proceed as described in Hudson, T.J. *et al.* (Science 270:1945-1954 (1995)) and Dietrich *et al.* (Dietrich, W. F. *et al.*, Nature 380:149-152 (1996); Dietrich, W. F. *et al.*, Nature Genetics 7:220-245; Dietrich, W. *et al.*, Genetics 131:423-447 (1992)). For example, in a 50 µl reaction volume, 0.5 ng of template nucleic acid/target polynucleotide is added

EXAMPLE 3

A set of tag sequences is selected such that the tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences. An oligonucleotide array of all of the tags is fabricated. The design and use of such a 4,000-20mer-tag array for the functional analysis of the yeast genome has been described (1). More recently, Affymetrix designed and fabricated an array with a set of more than 16,000 such tags. The tag sequence synthesized on the chip can be 20-mer, 25-mer, or other lengths.

EXAMPLE 4

Marker specific primers are used to amplify each genetic marker (*e.g.* SNP). A multiplex PCR strategy is used to amplify these markers from genomic DNAs of tested individuals (2). After PCR amplification, excess primers and dNTPs are removed enzymatically. These enzymatically treated PCR products then serve as templates in the next SBE reaction. Please note that these templates (PCR products) are double stranded, which are different from the templates used in other protocols (3, 4). For example, in Minisequencing (3) and Genetic Bit Analysis (GBA, 4), a double stranded template has to be converted to a single stranded template prior to the base extension reaction. The methods used for this conversion are costly, laborious, and hard to automate.

EXAMPLE 5

In the protocol described below, an SBE primer is designed for each genetic marker which terminates 1 base before the polymorphic site. However, other primer design schemes can be used. The primer for each marker is tailed with an unique tag which is complementary to a specific probe sequence synthesized on the tag chip. The extension reaction is multiplex, in which SBE primers corresponding to multiple markers were added in a single reaction tube, and extended in the presence of pairs of

10XPCR Buffer II	2.5 µl
25 mM $MgCl_2$	5 µl
25 mM dNTPs	1 µl
AmpliTaq Gold (5U/µl)	0.4 µl
ddH_2O	up to 25 µl

PCR conditions

96°C	10 min

40 cycles :

94°C	30 sec
57°C	40 sec
72°C	1 min 30 sec

72°C	10 min
4°C	O/N

Enzymatic treatment of PCR products to degrade and de-phosphorylate the unused primers and dNTPs, respectively:

To a 25 µl PCR products, add 1 µl of Exonuclease I (Amersham Life Science, 10 U/µl) and 1 µl of Shrimp Alkaline Phosphatase (Amersham Life Science, 1 U/µl), and incubate at 37° C for 1 hour. Inactivate the enzyme activities at 100°C for 15 minutes. Apply the sample to a S-300 column (Pharmacia), to further reduce the residual PCR primers and dNTPs, and replace the buffer with ddH_2O. The sample is ready for next SBE reaction.

Precipitation:

After SBE reaction, we combined 9 tubes for each sample, mix with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), then precipitated with 18.75 µ l of 8 M LiCl, and 1125 µl of pre-chilled (-20°C) ethanol (Abs.). Mix well; then centrifuge at the top speed (Eppendorf centrifuge 5415C) for 15 min at room temperature; Decant the supernatant, and dry the samples at 40C for 40 min, re-suspend the samples in 33 µl ddH2O, now it is ready for hybridization.

Hybridization:

The prepared sample is denatured at 100°C for 10 minutes and snap cooled on ice for 2-5 minutes. The universal tag chip is pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5mg/ml of BSA, then hybridized with 120 µl hybridization solution (as shown below) at 42°C 2 hours on a rotisserie, at≅ 40 RPM.

The hybridization solution contains:

5M TMACL	72 µl
0.5M MES (pH 6.7)	12 µl
1% Triton X-100	1.2 µl
HS DNA (10mg/ml)	1.2 µl
Flu-c213 (5 nM)	1.2 µl
BSA (20 mg/ml)	3.0 µl

Plus 29.4 µl prepared sample (see above).

Post-Hybridization Wash:

Rinse the chip with 1X SSPE-T 10" twice first, then wash with 1X SSPE-T for 15-20min at 40°C on a rotisserie, at ≅ 40 RPM. And then wash on a fluidic station (FS400, Affymetrix) 10 times with 6 x SSPET at 22°C.

The tag array strategy begins with an array of tag sequences selected in a manner that all tag probes are in the same length, *e.g.* 20-nucleotide long, with similar melting temperature and G-C content, and the lowest sequence homologous among each other[11]. Therefore, these tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences.

The design and use of a 4,000-tag array for the functional analysis of yeast *Saccharomyces cerevisiae* genes[11] and drug sensitivity studies[12] have been described. More recently, we have designed and fabricated an array that contains more than 32,000 such tags, and developed it as a genotyping tool, in combination with marker-specific PCR amplifications and SBE reactions.

As shown in Fig. 7, marker specific primers are designed and used to amplify each single nucleotide polymorphism (SNP). A multiplex PCR strategy is used to amplify these SNPs from genomic DNAs[9]. In general, SNPs with same base composition at the polymorphic site (*e.g.* all the A/G polymorphisms) are grouped together. After PCR amplification, excess primers and dNTPs are degraded and de-phosphorylated using Exonuclease I and Shrimp Alkaline Phosphatase, respectively. These enzymatically treated PCR products (double-stranded) are then served as templates in the SBE reaction. A SBE primer is designed for each genetic marker, which terminates one base before the polymorphic site. Each primer is tailed with a unique tag that is complementary to a specific probe sequence synthesized on the tag array. The extension reaction is multiplex, in which SBE primers corresponding to multiple markers (up to 56 markers that we have tested so far) were added in a single reaction tube, and extended in the presence of pairs of ddNTPs labeled with different fluorophores, *e.g.* for an A/G variant, biotin-labeled ddATP and fluorescein-labeled ddGTP are used. The resulting mixture of SBE reactions is hybridized to the tag array. Each tag hybridizes to a specific probe position on the chip. The ratio of the intensities of the colors indicates the genotype (homozygous wild type, or homozygous mutant, or heterozygous) or the allele frequency (ranging from 0% to 100%) in the samples tested.

multiplexing SBE assay was developed with a complexity of 9 to 28 markers in each reaction and a total of 9 reactions for the 165 markers. 21 of them (12.7%) failed in the multiplexing PCR and multiplexing SBE assay. Therefore, 144 markers from 49 genes passed the assay development. The gene location, polymorphic sites, and the designed primers for these 144 markers were summarized in Table 1.

We then genotyped 44 individuals using 44 tag arrays. Good hybridization signals were obtained in 96.5% (6116 / 6336 (144 x 44)) of the cases. The signal intensity values from the hybridization results were used in clustering analysis for each of the 144 markers. Genotypes for each individual at the 144 loci were assigned automatically based on the clustering results, with some manual editing. Data Desk 6.0 (Data description, Inc.) was used to manually display the clustering analysis results (of the intensity ratios of the two colors). Overall, 80-85% of the markers form good cluster(s).

We have performed the gel-based DNA sequencing to determine the genotypes at 115 loci in 3 of the 44 individuals (see Methods). Comparison of the ABI sequencing results and the chip results resulted in 14 discrepancies (4%), out of 115 x 3 = 345 genotype calls. Most of the discrepancies occurred in cases where one method called homozygous, while the other method called heterozygous. In one case (marker ICAM1ex6.254), where the ABI sequencing method called G/G, but the tag array /SBE assay method called A/A in all the three individuals, we believe the discrepancies are due to mis-priming of the SBE primer to adjacent sequences.

We also tested the reproducibility of the tag array/SBE assay genotyping method. We repeated the multiplexing PCR, SBE and the chip hybridization experiments in 4 individuals. The ratios of the two colors (for each of the 144 markers) in the replicated experiments are not all exactly the same, but they all fall into the same cluster (i.e. giving the same genotype call). Therefore, we didn't find any discrepancy in the genotyping call of duplicated samples.

ACEEX21.150	CATGAGGCCA(T/C)TGGGGACGTG	CGGCTTCCATGAGGCC	GGCTAGCACGTCCCCAA	GATCTGGCTTCAACT GTATGCCGGCTTCCA TGAGGCCA
ACEEX22.19	TGACATCAAC(T/G)TTCTGATGAA	TTGCAGAGCATGACATCAA	AAGGGCCATCTTCATCAGA	TGCCTAGCTTTCCATA TCGGCCTTGCAGAGC ATGACATCAAC
ACEEX24.118	CCAAGGAGGC(C/T)GGGCAGCGCC	CATCTACCAGTCCAAGGAG G	TCACCCAGGCGCTGC	TATCTCGCTTGCTATC AACGATCTACCAGTC CAAGGAGGC
ACEEX24.16	CCAGGTACTT(T/C)GTCAGCTTCA	TCGCTCTGCTCCAGGTACT	GGAACTGGATGATGAAGCT GA	GCCTAAGCTCTGTCG CTGATTCGCTCTGCTC CAGGTACTT
ACEEX26.154	CTCAGCCAGC(G/A)GCTCTTCAGC	CTGGGCCTCAGCCAG	GCGGATGCTGAAGAGCC	TCTATTGCTGTTCGGC GGCAACCCTGGGCCT CAGCCAGC
ACEEX26.174	CATCCGCCAC(C/A)GCAGCCTCCA	TCTTCAGCATCCGCCA	GCCGGTGGAGGCTGC	AGCAGAGATGGACAG ACCTCCTCTTCAGCAT CCGCCAC
ACEEX26.205	CACGGGCCCC(A/C)GTTCGGCTCC	CACTCCCACGGGCCC	CACCTCGGAGCCGAACT	GCTGGCGGTTCATGC AATCTTCCACCTCGG AGCCGAAC
ACEEX26.224	CCGAGGTGGA(G/A)CTGAGACACT	TCGGCTCCGAGGTGG	CACCTCAGGAGTGTCTCAGC	TATCTGCGTTGCTGAC GTGCCAGTTCGGCTC CGAGGTGGA
ACEEX8.106	AGGATCTGCC(C/T)GTCTCCCTGC	CCTGCAGTACAAGGATCTG C	CCCGACGCAGGGAGA	GATCCGTATGTCGAA TGGCTCTGCAGTACA AGGATCTGCC
ACEP.-3892	TAAGGGGGGG(T/C)TGCTGTACAT	CCACTGAGGATAAGGGGG	GAAGATATTTGCAAAGTAT GTACAGC	CCAGAGGTGCGGTCA CATATCACTGAGGAT AAGGGGGGG

AGTEX2.354	GGATGCTGGC(C/T)AACTTCTTGG	TGGTCGGGATGCTGG	CGGAAGCCCAAGAAGTTG	TTTCGTGCTTTGGAG ACAGCAATGGTCGGG ATGCTGGC
AGTEX2.755	TTCACAGAAC(T/G)GGATGTTGCT	CGCTCTCTGGACTTCACAGA	TCTCAGCAGCAACATCCA	TGCCGTGTTGGTGCTT CACACTCTCTGGACTT CACAGAAC
AGTEX2.827	TGCTCCCTGA(T/C)GGGAGCCAGT	AGACTGGCTGCTCCCTG	TCCACACTGGCTCCCA	TCGTCCACTTTAGCAT GATGAAGACTGGCTG CTCCCTGA
AGTEX5.376	GGAAAGCAGC(C/G)GTTTCTCCTT	GACTTTGAGCTGGAAAGCA G	CATGCAGCACACTTAGACC A	TACATACTTGCAGTG CGTTCACTTTGAGCTG GAAAGCAGC
AGTEX5.385	CCGTTTCTCC(T/C)TGGTCTAAGT	GACTTTGAGCTGGAAAGCA G	CATGCAGCACACTTAGACC A	CGTCGTGCTGCGTGA CTATAGGAAAGCAGC CGTTTCTCC
AGTEX5.641	TCGGTTTGTA(T/G)TTAGTGTCTT	GCATTGCCTTCGGTTTGT	TCATGTTCTTACATTCAAGA CACTAAA	TGAGAGTCTGTTCTT AGGCCCATTTTTGCAT TGCCTTCGGTTTGTA
AGTEXP1.101	CTGTGCTATT(G/C)TTGGTGTTTA	CTTTCAATCTGGCTGTGCTA T	GGGGAGACTGTTAAACACC AA	TACATAATTGCCATG ACGGGTTCAATCTGG CTGTGCTATT
AGTEXP2.160	CCTTGGCCCC(G/A)ACTCCTGCAA	TGGGAACCTTGGCCC	ACCGAAGTTTGCAGGAGTC	GAGAATGCTGTATAG TGTCCTTTCTGGGAAC CTTGGCCCC
AGTEXP2.203	ACCCTGCACC(G/A)GCTCACTCTG	TGTGTAACTCGACCCTGCAC	CTGCTGAACAGAGTGAGCC	CGTCTCGCTGGTCACT AATGGTGTAACTCGA CCCTGCACC
AGTEXP2.35	CTGCACCTCC(G/A)GCCTGCATGT	TCTGCCCTCTGCACCTC	CAGGGACATGCAGGCC	GATCTCTGTGAAGTT AGTGCCCTCTGCCCTC TGCACCTCC

APOC2.804	CTTTCTCCCC(A/T)GGGACTTGTA	ACCATCTGTGCTTTCTCCC	TCATGGCTGCTGTGCTT	TGAGAAGTTGTGAAG ATCCCTAACCATCTGT GCTTTCTCCCC
APOC2.819	CTTGTACAGC(A/C)AAAGCACAGC	ACCATCTGTGCTTTCTCCC	TCATGGCTGCTGTGCTT	GCCAGGCGTTCAGAT GCAATCCCAGGGACT TGTACAGC
APOC4.3162	CTGGGTCCGC(T/G)CACCAAGGCC	AGGGACCTGGGTCCG	AGGAACCAGGCCTTGGT	GCTGGTCGTGGTCCA ATCATTGAGGGACCT GGGTCCGC
APOER2EX12.68	ACTGTCCAGC(A/C)TTGACTTCAG	CAAGCTACACCAACTGTCC AG	TCTGTTGCCTCCACTGAAG	GACCATGCTGGCTTA CCTGTAAGCTACACC AACTGTCCAGC

CHYEX2.168	ACGGCTGCTC(A/G)TTGTGCAGGA	TGTGCTGACGGCTGCT	TGTCTCACCTTCCTGCACA	CCATCGAATCGTCTA TCAGTACTTTGTGCTG ACGGCTGCTC
CLCNKBEX10.33	GGCCACCTTG(G/C)TTCTCGCCTC	CCGCTCTGGCCACCTT	AGGTGATGGAGGCGAGA	GGTCTCAATTAGGCT TCATGTACTCCGCTCT GGCCACCTTG
CLCNKBEX15.64	GCCAAGGACA(C/T)GCCACTGGAG	CCACACTGGCCAAGGA	CCTTGACCACCTCCTCCA	GCCGGTCATGTGCTC TGATATCACCACACT GGCCAAGGACA
CLCNKBEX4.19	AATCCCGGAG(G/C)TGAAGACCAT	GGTTCTGGAATCCCGGA	CCGCCAACATGGTCTTC	GCGTGATATTCCATG ATCTGAGGTTCTGGA ATCCCGGAG
CLCNKBEX4.70	GGATATCAAG(A/C)ACTTTGGGGC	TGGAGGACTACCTGGATAT CAA	CCACTTTGGCCCCAAA	GCTGGTGATGGCTCT TCATATGGAGGACTA CCTGGATATCAAG
COX2EX1.358	CCAATTGTCA(T/G)ACGACTTGCA	CGGTTAGCGACCAATTGTC	GACGCTCACTGCAAGTCG	CGAACATCTGTCACA ATGCGCTCGGTTAGC GACCAATTGTCA
COX2EX10.156	ATGGTAGAAG(T/C)TGGAGCACCA	TTTGGTGAAACCATGGTAG AA	TCAAGGAGAATGGTGCTCC	GACTCTAGTGTCGTCT GATCTCTTTGGTGAA ACCATGGTAGAAG
CYP11BIEX4.205	AGGAGCACTT(T/G)GAGGCCTGGG	AAGGTGTGGAAGGAGCACT	ATGCAGTCCCAGGCCT	TCAGATGTTGTAATC GTGCGCAAGGTGTGG AAGGAGCACTT
CYP11BIEX5.107	CGTGGCGGAG(C/G)TCCTGTTGAA	CAGTACACCAGCATCGTGG	AGTTCCGCATTCAACAGG	GCGTCGGCTTCATGC GATATTACACCAGCA TCGTGGCGGAG
CYP11B2EX3.152	CAGGCCCTGA(A/G)GAAGAAGGTG	GCAGTGGCCAGGGACT	CGTTCTGCAGCACCTTCTT	ATGCACGATCCTCTA CATTGGGACTTCTCCC AGGCCCTGA

eNOS.78	CCCCAGATGA(T/G)CCCCCAGAAC	TGCAGGCCCCAGATG	CAGAAGGAAGAGTTCTGGG G	ATACGGGATGATGAG CATACTGCTGCAGGC CCCAGATGA
ET1EX5.90	TGAAAGGCAA(G/T)CCCTCCAGAG	TCCCAAGCTGAAAGGCA	CACATAACGCTCTCTGGAGG	TACATGACTTGCCCT GCTGTTTCATGATCCC AAGCTGAAAGGCAA
GALNREX1.327	GCACGCAGCC(G/C)CTCCGGGAGC	CAGGTGCAGCACGCA	TCCCTGGCTCCCGGA	ACGATGAGCAGGGAT CACTAACAGGTGCAG CACGCAGCC
GALNREX1.553	TCAGAAGGTC(G/C)CGGCGCAAAG	CCCACCCTCTCTCAGAAGGT	CACCGTCTTTGCGCC	ATCTGAGAGCTAGTC GGCATCCACCCTCTCT CAGAAGGTC
GGREX9.29	AACATGGGCT(T/G)CTGGTGGATC	AGCAATGACAACATGGGC	CCGCAGGATCCACCA	GGTGACTATTCGGCT GCTCTACCAGCAATG ACAACATGGGCT
GLUT2EX1.137	AGCACTAATT(C/A)TCTGTGGAGC	CTAAACAGAAACACCACAG CAC	ACTGCACTCTGCTCCACAG	TAGCTGTGTTGACAT CTGGCACAGAAACAC CACAGCACTAATT
GLUT2EX1.164	CAGTGTGCCT(T/C)CCATGCTCCA	GCAGAGTGCAGTGTGCC	GCTGTGCTGTGGAGCATG	TGCTTAGTTGTGAGT CGCCAGAGCAGAGTG CAGTGTGCCT
GLUT4EX3.112	GGCACCCTCA(C/G)CACCCTCTGG	CCCTCCAGGCACCCTC	AGAGGGCCCAGAGGGT	CTCACGACTGGGCTG ATGATTCCATCCCTCC AGGCACCCTCA
GMP-140.105	TTTCTCTTGT(A/G)ACAATGGCTT	TGGAGCGGTGGCTTCTA	CCCACCCATTATCAGACCTA	TGGCACAGTTTCCTG CTGGTGGCTCCACCT GTCATTTCTCTTGT
GMP-140.164	CCACTGGTCA(A/C)CTACCGTGCC	AAGAGAATGGCCACTGGTC	GCAGGTTGGCACGGTA	GCTGGGTGTGATCCT CTCTACAAGAGAATG GCCACTGGTCA

ICAM1EX6.254	GGTCACCCGC(G/A)AGGTGACCGT	AGGGGAGGTCACCCG	AGCACATTCACGGTCACC	GTGCTGGGTTCGCAT TCATCGCACATTCAC GGTCACCT
ICAM1EX6.39	GATGGCCCCC(G/A)ACTGGACGAG	TTTTCCAGATGGCCCC	GACAATCCCTCTCGTCCAG	CCAATAGGTGCTCAC GTCATGTGTTTTTCCA GATGGCCCCC
ICAM2EX2.63	AAAGAAGCTG(G/A)CGGTTGAGCC	CGTGAGGCCAAAGAAGCT	CCCTTTGGGCTCAACC	TTGGCTCATTTGCATG GCGCCACGTGAGGCC AAAGAAGCTG
IRS-2.AA1057	CAGGGCCCGG(G/A)CGCCGCCTCA	GTTGCCACCGCCCAG	CAACGATGAGGCGGC	TGCTCGCTTGTGATCG ACTGTTGCCACCGCC CAGGGCCCGG

NETEX5.121	AATGGCATCA(A/C)TGCCTACCTG	GAGCCTCCAATGGCATC	GTCGATGTGCAGGTAGGC	GCATGAAGTTCCATA ATCGCGAGCCTCCAA TGGCATCA
NETEX7.112	TGGTTACATG(G/C)CCCATGAACA	TCTTCTCCATCCTTGGTTAC A	TGTTGACCTTGTGTTCATGG	CAGTGACATGCCGCT CAGTACATCTTCTCCA TCCTTGGTTACATG
NETEX7.131	CACAAGGTCA(A/G)CATTGAGGAT	GCCCATGAACACAAGGTC	TGTGGCCACATCCTCAAT	CGGCAATATGATGAT AGGTCCCCATGAACA CAAGGTCA
NETEX7.73	CACCAGCTTC(G/C)TCTCTGGGTT	GCATCAACTGTATCACCAGC TT	AGATGGCGAACCCAGAG	CCTGGTATGACATGG AGCCTCAGCATCAAC TGTATCACCAGCTTC
NETEX9.157	TGCATAACCA(A/G)GGTGAGTAGG	CGCCCTGTTCTGCATAACC	GCCCAGCCCCTACTCAC	CCAACGATGCTACTG AGTCACGCCCCTGTTCT GCATAACCA
OB.160	GATCAATGAC(A/G)TTTCACACAC	AATTGTCACCAGGATCAAT GA	ACTCTCCTTACCGTGTGTGA A	CATTGCACCCACTGA GATGGATTGTCACCA GGATCAATGAC
OB-R.174	GTAATTTTCC(A/G)GTCACCTCTA	TCACATCTGGTGGAGTAATT TTC	GCTGAACTGACATTAGAGG TGA	CACGGATCTGCCGCT AGAATCATCTGGTGG AGTAATTTTCC
PG1SEX1.396	GGGAGCAGGG(T/G)TTCTCCCAGA	GCTGCGGGGAGCAGG	GGGCGCTCTGGGAGA	CGAACACATGCGGCT GGATAAGCTGCGGGG AGCAGGG
PLA2AEX2.42	GCCGCCGCCG(A/C)CAGCGGCATC	CTTGCAGTGGCCGCC	AGGGCTGATGCCGCT	AGATAGAGTCGATGC CAGCTTTGCAGTGGC CGCCGCCG
PLA2AEX3.104	TGCTGGACAA(C/A)CCGTACACCC	TGGACAGCTGTAAATTTCTG CT	ATGAATAGGTGTGGGTGTA CG	TGCCTCATTGTGACTC ATGGACAGCTGTAAA TTTCTGCTGGACAA

SCNN1GEX1.236	GTCGTGGCCC(G/T)CTCCGGGCGG	CGTTGTGAAGTCGTGGCC	CTGAGACCGCCCGGA	CAGTGACGTGAGTGC CATCTGTTGTGAAGT CGTGGCCC
SCNN1GEX2.219	GGTGTCCCGC(G/T)GCCGTCTGCG	GCATCGTGGTGTCCCG	GGAGGCGGCGCAGAC	CTCAGCAGTTAGCAG CGCATCGCATCGTGG TGTCCCGC
SCNN1GEX3.259	GCGGAAAGTC(G/A)GCGGTAGCAT	GGGAGGAAGCGGAAAGT	GAAGCCTTGTGAATGATGCT	CTTATGGCGCTGTCG GCTATCAGGGAGGAA GCGGAAAGTC
TBXASEX11.88	CCCCGCAGGC(G/A)CTGTGCTAGA	CGAGGTGCTGGGGCA	ACGGCCATCTCTAGCACA	GATATGCGTTACGTG AGTCTCGGCCATCTCT AGCACAG
TBXASEX9.276	TGCCACCTAC(C/G)TACTGGCCAC	CACACTTTCTTTTGCCACCT	AGGGTTGGTGGCCAGT	CAACAACTGCGCGCGAC GATGAAACACACTTT CTTTTGCCACCTAC
TGF-B1.75	CTCATGGCCA(C/T)CCCGCTGGAG	TCCTGCTTCTCATGGCC	GGCCCTCTCCAGCGG	TTGTGCATTGTTGGA CGCCCCTTTCCTGCTT CTCATGGCCA
TRHREX1.56	GCAGAACTTA(G/C)ATGATAAGCA	CAGGTACTAGAGTTTCTGCA GAACTT	GGCTTTGTCGTTGCTTATCA	AGCAGTAATGACAGC GTGCAAGGTACTAGA GTTTCTGCAGAACTT A

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex;

(c) subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs;

(d) contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementarysequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags,

thereby determining the genotype of the queried nucleotide in the sample.

4. A method to aid in determining a ratio of alleles at a polymorphic locus in a sample, comprising the steps of:

(a) using a pair of primers to amplify a region of a nucleic acid in a sample, wherein the region comprises a polymorphic locus, whereby an amplified DNA product is formed;

(b) labeling an extension primer by a single base extension reaction to form a labeled extension primer, wherein the amplified DNA product is used as a template, wherein the extension primer comprises a 3' portion and a

13. The method of claim 4 wherein the step of labeling employs at least two distinct dideoxynucleotides bearing distinct labels.

14. The method of claim 4 wherein the step of labeling employs four distinct dideoxynucleotides bearing distinct labels.

15. The method of claim 4 further comprising the steps of:

(d) comparing quantities of a first and a second label at a location on the solid support; and

(e) determining the ratio of nucleotides present at the polymorphic locus in the sample.

16. The method of claim 15 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

17. The method of claim 4 wherein the sample comprises DNA from two or more individuals.

18. The method of claim 17 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

19. The method of claim 4 wherein the solid support is selected from the group consisting of beads, microtiter plates, and oligonucleotide arrays.

20. A set of primers for use in determining a ratio of nucleotides present at a polymorphic locus, comprising:

(a) a pair of primers which when in the presence of a DNA polymerase amplify a region of double stranded DNA. wherein the region comprises a polymorphic locus; and

dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer, wherein each type of dideoxynucleotide present in the reaction bears a distinct label; and

(b) hybridizing the 5' portion of the extension primer to one or more probes complementary to the 5' portion which are immobilized to known locations on a solid support.

27. The method of claim 26 wherein two complementary strands of the DNA molecule are present in the single base extension reaction..

28. The method of claim 27 wherein each complementary strand of the DNA molecule is used as a template to label an extension primer.

29. The method of claim 26 wherein the label is a fluorescent label.

30. The method of claim 26 wherein the label is a radiolabel.

31. The method of claim 26 wherein the label is an enzyme label.

32. The method of claim 26 wherein the label is an antigenic label.

33. The method of claim 26 wherein the label is an affinity binding partner.

34. The method of claim 26 further comprising the step of:

(c) optically detecting a fluorescent label on the solid support.

35. The method of claim 26 further comprising the steps of:

(c) comparing quantities of a first and a second label at a location on the solid support; and

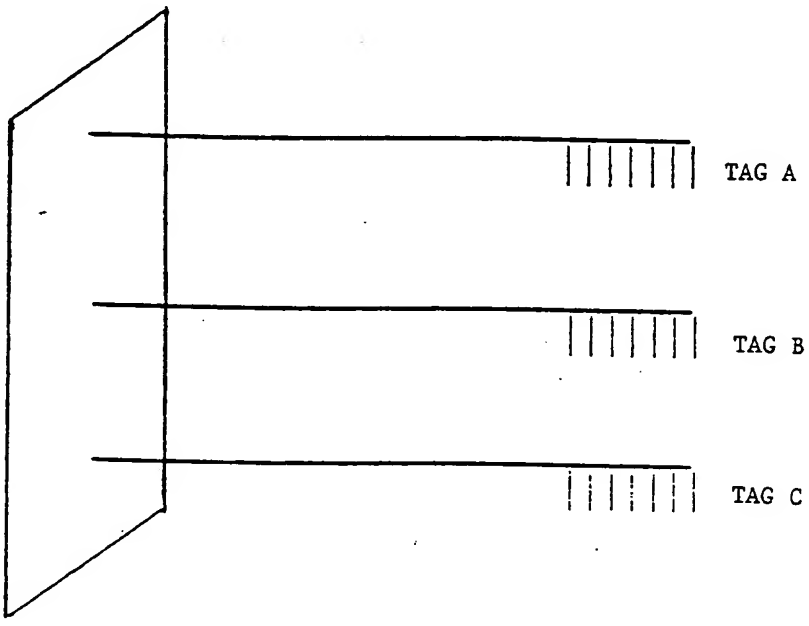
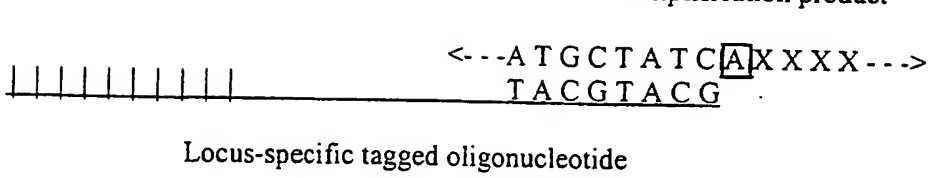


Fig. 1



Locus-specific tagged oligonucleotide

Fig. 3

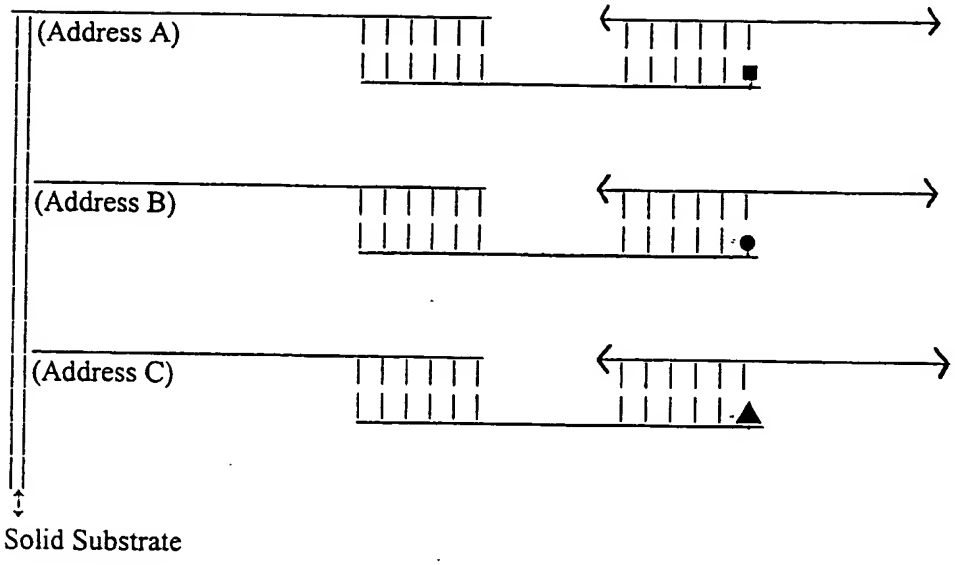


Fig. 5

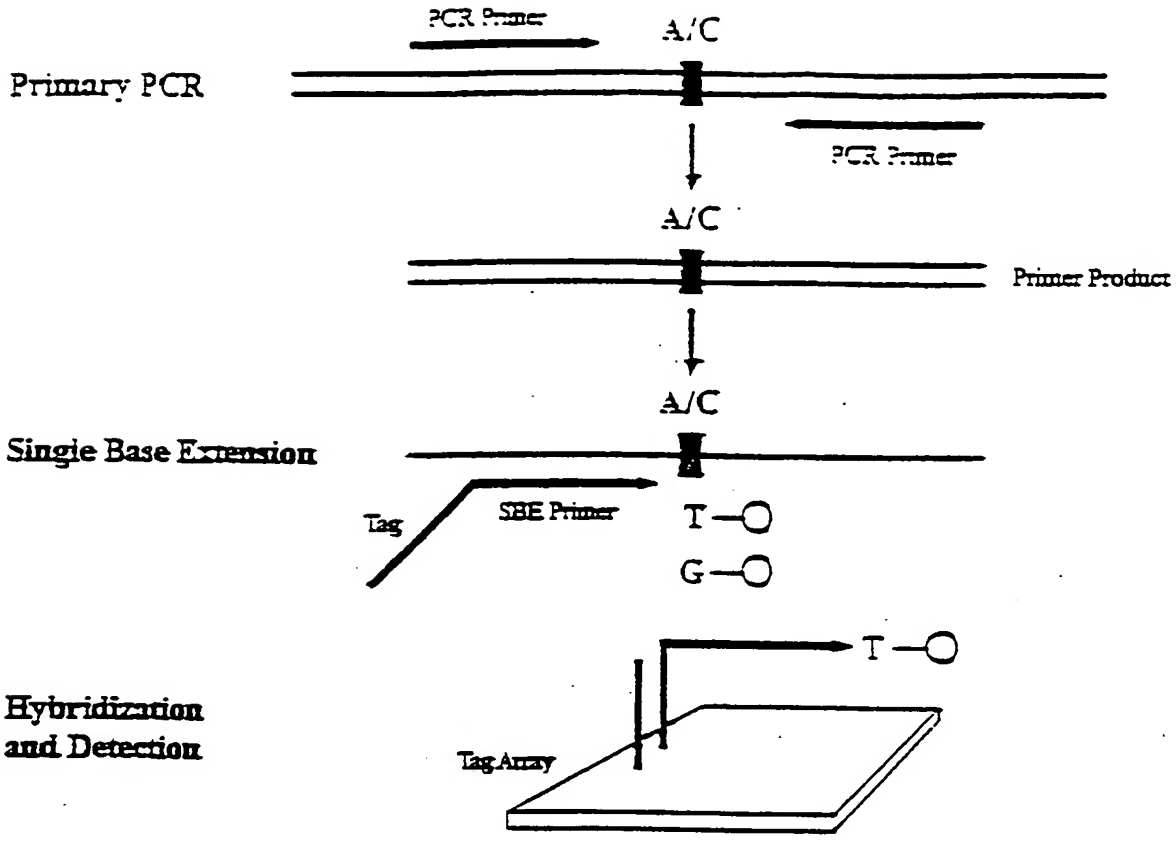


Fig. 7

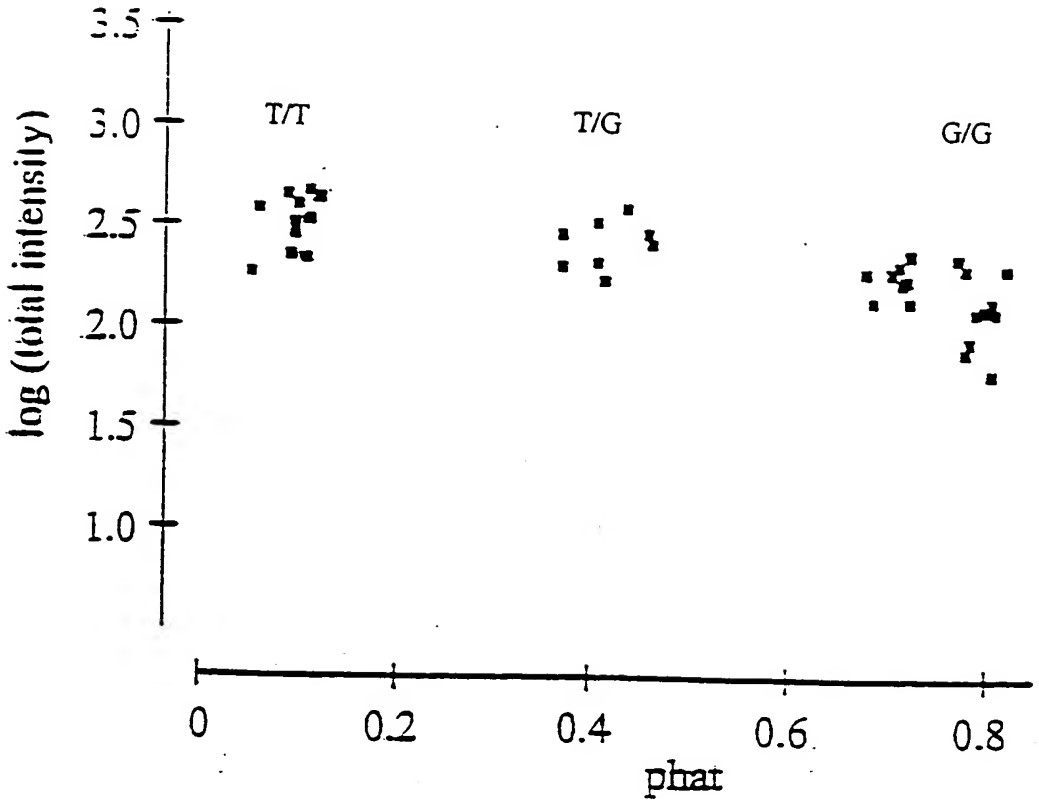


Fig. 9

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 October 2000 (05.10.2000)

PCT

(10) International Publication Number
WO 00/58516 A3

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US00/08069

(22) International Filing Date: 27 March 2000 (27.03.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/126,473 26 March 1999 (26.03.1999) US
60/140,359 23 June 1999 (23.06.1999) US

(71) Applicants: WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors: FAN, Jian-Bing; 3737 Nobel Drive, #2115, San Diego, CA 92122 (US). HIRSCHHORN, Joel, N.; 14 Oakmont Road, Newton, MA 02459 (US). HUANG, Xiaohua; 937 Jackson Street, Mountain View, CA 94043 (US). KAPLAN, Paul; 2190 Vizcaya Circle, Campbell, CA 95008 (US). LANDER, Eric, S.; 151 Bishop Allan Drive, Cambridge, MA 02139 (US). LOCKHART, David, J.; 510 Torrey Point Road, Del Mar, CA 92014 (US). RYDER, Thomas; 100 Spring Street, Los Gatos, CA 95030 (US). SKLAR, Pamela; 68 Fuller Street, Brookline, MA 02446 (US).

(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

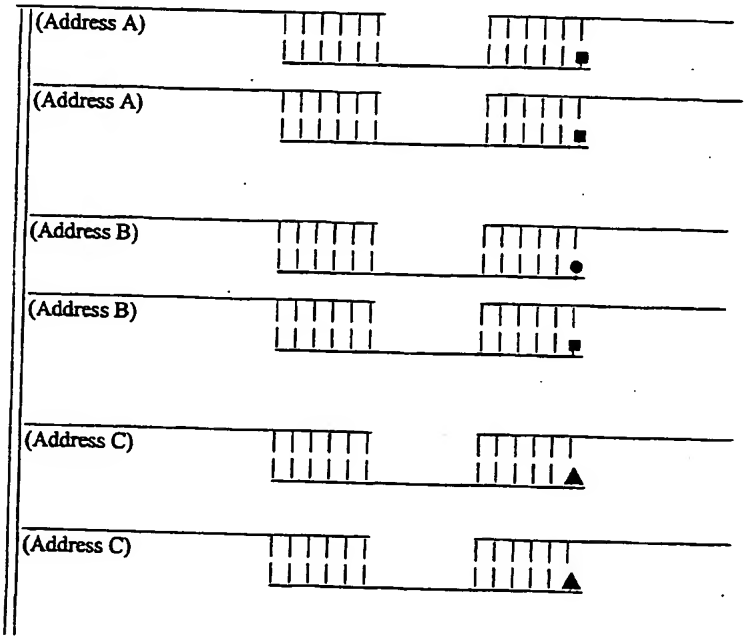
(81) Designated States *(national)*: CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *with international search report*

[Continued on next page]

(54) Title: UNIVERSAL ARRAYS



(57) Abstract: An array of oligonucleotides on a solid substrate is disclosed, which can be used for multiple purposes. Methods and reagents are provided for performing genotyping to determine the identity or ration of allelic forms of a gene in a sample. A single base extension primer is coupled to a sequence identity code. During the primer extension reaction a distinctive label is incorporated which identifies the allelic form present in the sample. This permits multiple simultaneous analyses to be performed easily and efficiently.

WO 00/58516 A3

INTERNATIONAL SEARCH REPORT

Inten nal Application No
PCT/US 00/08069

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, MEDLINE, CHEM ABS Data, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 93 25563 A (HOPE CITY ;WALLACE ROBERT BRUCE (US)) 23 December 1993 (1993-12-23) the whole document	1-40
X	WO 97 31256 A (BLOK HERMAN ;BARANY GEORGE (US); KEMPE MARIA (US); ZIRVI MONIB (US) 28 August 1997 (1997-08-28)	1,2
Y	the whole document	3-40
Y	PASTINEN T ET AL: "MINISEQUENCING: A SPECIFIC TOOL FOR DNA ANALYSIS AND DIAGNOSTICS ONOLIGONUCLEOTIDE ARRAYS" GENOME RESEARCH,US,COLD SPRING HARBOR LABORATORY PRESS, vol. 7, no. 6, 1 June 1997 (1997-06-01), pages 606-614, XP000699761 ISSN: 1088-9051 the whole document	3-40

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 January 2001	05/02/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hagenmaier, S

Form PCT/ISA/210 (second sheet) (July 1992)

1

INSDOCID: <WO___0058516A3_I_>

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inten nal Application No
PCT/US 00/08069

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9325563 A	23-12-1993	CA 2115342 A	23-12-1993
		AU 674127 B	12-12-1996
		AU 2251192 A	04-01-1994
		EP 0607151 A	27-07-1994
		JP 6509946 T	10-11-1994
		US 5981176 A	09-11-1999
WO 9731256 A	28-08-1997	AU 2799797 A	10-09-1997
		CA 2244891 A	28-08-1997
		EP 0920440 A	09-06-1999
EP 0799897 A	08-10-1997	NONE	
EP 0416817 A	13-03-1991	AT 144290 T	15-11-1996
		AU 622349 B	02-04-1992
		AU 6118490 A	14-03-1991
		DE 69028891 D	21-11-1996
		DE 69028891 T	13-03-1997
		GB 2235689 A,B	13-03-1991
		JP 3272686 A	04-12-1991
		NZ 235015 A	26-03-1992
		US 5525494 A	11-06-1996
GB 2312747 A	05-11-1997	AU 2704997 A	26-11-1997
		EP 0896630 A	17-02-1999
		WO 9742345 A	13-11-1997
		JP 2000510337 T	15-08-2000
US 5710028 A	20-01-1998	IL 102382 A	04-08-1996
		IL 111267 A	14-07-1999
		AU 4770393 A	31-01-1994
		EP 0648222 A	19-04-1995
		FI 946127 A	20-02-1995
		IL 106199 A	24-09-1998
		JP 8507198 T	06-08-1996
		NO 945100 A	22-02-1995
		WO 9401447 A	20-01-1994
		US 5650277 A	22-07-1997
WO 9641011 A	19-12-1996	US 5635400 A	03-06-1997
		AU 718357 B	13-04-2000
		AU 6102096 A	30-12-1996
		CA 2222581 A	19-12-1996
		CN 1193357 A	16-09-1998
		CZ 9703926 A	17-06-1998
		EP 0832287 A	01-04-1998
		HU 9900910 A	28-07-1999
		JP 11507528 T	06-07-1999
		NO 975744 A	05-02-1998
		PL 324000 A	27-04-1998
		AU 712929 B	18-11-1999
		AU 4277896 A	06-05-1996
		AU 7717596 A	30-04-1997
		CZ 9700866 A	17-09-1997
		EP 0793718 A	10-09-1997
		EP 0931165 A	28-07-1999
		FI 971473 A	04-06-1997
		JP 10507357 T	21-07-1998